

DOCUMENT RESUME

ED 355 769

FL 021 004

AUTHOR Lewkowicz, Jo A.
TITLE Assessing Students at Tertiary Level: How Can We Improve?
REPORT NO ISSN-1015-2059
PUB DATE 92
NOTE 10p.; For the serial issue from which this paper is analyzed, see FL 021 000.
PUB TYPE Viewpoints (Opinion/Position Papers, Essays, etc.) (120) -- Journal Articles (080)
JOURNAL CIT Hongkong Papers in Linguistics and Language Teaching; v15 p39-46 1992

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Educational Assessment; *Evaluation Criteria; *Evaluation Methods; Foreign Countries: Higher Education; *Student Evaluation; Workshops
IDENTIFIERS University of Hong Kong

ABSTRACT

A number of issues related to assessment of students undergoing English enhancement courses were raised at a workshop on assessment held at The University of Hong Kong Language Centre. The primary focus of the workshop was to update staff about current assessment practices in the various programs run by the Language Centre and to discuss issues of professional interest. The workshop threw light on some of the persistent problems in assessment that are experienced by a rapidly expanding tertiary teaching program and should be helpful to others facing a similar situation. Questions covered relate to types of tests used, who assessment is for, authenticity, assessment criteria, and alternative assessment. Although many problems remain unresolved, the exchange of ideas reported has suggested lines for future investigation and development. Contains 7 references (LB)

* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *

Assessing Students at Tertiary Level: How Can We Improve?

Jo A. Lewkowicz

This paper develops a number of issues, related to assessment of students undergoing English enhancement courses, raised at a workshop on assessment held at the Language Centre¹ of HKU. The primary focus of the workshop was to update staff about current assessment practices in the various programmes run by the Language Centre and to discuss issues of professional interest. The workshop threw light on some of the persistent problems in assessment that are experienced by a rapidly expanding tertiary teaching programme and should prove illuminating to others facing a similar situation. Although many problems remained unresolved, the healthy exchange of ideas reported has suggested lines for future investigation and development.

Introduction

Assessment is a pivotal activity in any teaching operation and it is essential that teachers within an institution are informed of the methods used and their underlying rationale. Such a process not only helps ensure a code of practice but also affords a starting point from which change and development can take place. As an institution grows and the number of students and teachers increase, there is likely to be a healthy divergence of views as to the functions and best modes of assessing students, yet it is important that assessment remains meaningful to the student and does not become idiosyncratic. It is equally important that change is allowed to evolve and that such evolution is a result of extensive discussion, piloting and evaluation as well as careful scrutiny of the methods used to effect the change.

The Situation

The English Section of the Language Centre is responsible for a number of courses across different faculties and the importance given to formal assessment depends largely on the accountability of the Language Centre to the department or faculty it is assisting. This in turn is a function of the percentage of first year students taught on the English courses and the relative importance of each course in relation to other faculty-based courses, i.e. whether the results of the English course are noted on the student's transcript and whether or not this course is credit-bearing. Table 1 summarizes present practice. Notice as one reads across the courses from left to right, how the degree of accountability to faculty increases. In the Medical Faculty the weakest 15% of the first year undergraduate students are required to take an English course and to satisfy Language Centre criteria, whereas in the Engineering Faculty all students of the Information Stream of Computer Science are required to take such a course and it counts as one full paper in their degree programme.

The assessment procedures in place have all been subjected to systematic and principled development and modification. Course designers as well as a small group of testers within the Centre take an active role in writing language tests, moderating and piloting them in preparation for reviewing students' performance. It would therefore appear that the present situation is satisfactory and does not require immediate change. However, if the Centre is to ensure that its assessment procedures are appropriate for its students, as well as informative and cost-effective, then it must continue to recognise that review needs to be built into the system.

Jo A. Lewkowicz is a lecturer in the English Centre of the University of Hong Kong where she has been since 1989. Prior to that she worked on a testing project in China. Her current area of research is in the authenticity of integrated language tests. She has taught in Europe, Africa and the Middle East.

"PERMISSION TO REPRODUCE THIS
MATERIAL HAS BEEN GRANTED BY

Jo A. Lewkowicz
Allison

TO THE EDUCATIONAL RESOURCES
INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)

☒ This document has been reproduced as
received from the person or organization
originating it
☐ Minor changes have been made to improve
reproduction quality

• Points of view or opinions stated in this docu-
ment do not necessarily represent official
OERI position or policy

Table 1: An Overview of the Current Assessment¹ Procedures 1991/92

	Surveying/ Architecture	Medics	Arts ²	Engineering	Science	Social Sciences	Business Communication
Pre-course tests	nil	oral: seminar (Selection)	<ul style="list-style-type: none"> writing (Selection) oral: seminar (Course evaluation & pedagogy) 	<ul style="list-style-type: none"> report writing word completion proofreading (Selection) 	<ul style="list-style-type: none"> writing listening reading (Selection) 	nil	nil ³
Continuous Assessment	nil	oral presentation	oral, writing projects	report writing	writing, projects	oral, writing, projects	projects
Post-course tests	nil	oral: seminar (Achievement)	<ul style="list-style-type: none"> integrated: reading, listening & writing (Achievement) oral: seminar (Course evaluation & pedagogy) 	report writing (Achievement)	writing, listening (Achievement)	integrated: reading, listening & writing (Achievement)	<ul style="list-style-type: none"> language usage written communication concepts of communication (Achievement)

Notes: (1) "Assessment" in this context refers to assessment which is systematically recorded in such official records as Language Centre student cards, Faculty transcripts or reports to Faculty. (2) A mid-course test is administered in December to selected Arts students for release purpose. (3) An oral test assessing information transfer skills is administered in the first lesson for diagnostic purposes.

What sort of Assessment?

Any reflection on current assessment procedure and practice first needed to consider:

- what type of assessment is most appropriate for the courses being run? and
- should there be any changes to existing practice as English enhancement courses are extended across all faculties of the university?

Two extreme scenarios were considered. The first was relying totally on an end-of-course test for student assessment whereas the second was assessing student performance entirely on course work through continuous assessment. After detailed discussion both these extreme views were rejected since it was recognised that tests and continuous assessment fulfil a different function and both contribute to the building up of student profiles.

Tests

Such summative measures are formal, standardized procedures that provide more objective information about students' performance. Although students may be given their grade or percentage mark for their performance, they do not receive detailed feedback and the focus for course evaluation purposes is on how well (or poorly) the group has performed rather than on the individual student. Hence the test information may be crucial for faculty and the administration of the Language Centre, but of more limited value to the student. Students at the end of any course will ultimately be interested in whether they have passed or failed, while the Language Centre may, and frequently does, want to assess gain over time, which is a relevant consideration for course evaluation.

Continuous Assessment

With no continuous assessment, too much emphasis would be put on the test. The teaching would be affected and students would have little incentive to work consistently throughout the year. The same would most probably hold true if the continuous assessment were not graded and students knew it did not count towards their final assessment.

However, for continuous assessment, grades should be secondary to the feedback given to students. If students are to be motivated they need to know what they have done well and where they have failed to achieve. They also need to know what objectives they should be striving towards, and here lies a fundamental weakness in many assessment systems. It is too often assumed that students know what the assessor is looking for and what criteria will be used for assessment purposes. Withholding this information may be a result of it not being systematized and readily available in a form that would be comprehensible to the students. But it is necessary information not only for the students but also for staff, especially in a situation like the Language Centre of the University of Hong Kong (HKU) where a large number of teachers are teaching the same courses and it is desirable for them to be using the same criteria for assessing course work.

Accepting the principle of sharing assessment criteria with students, be it for tests or continuous assessment, has serious implications for course design. The criteria have to be explicit. In addition, teachers have to be ready to demystify these during the course of their teaching and to adhere to them once they have been set in place. This means that a time lag has to be built in before any further changes or developments can take place.

Recent research by Alderson and his colleagues (reported in Alderson, 1991a) into test method has indicated that even among 'experts' there may be little agreement as to what a test is actually testing or to the difficulty of test items or tasks. It is therefore likely that marking criteria are subject to similar variations of interpretation which would suggest that teachers may need to be 'trained' in the use of such criteria if reliability is to be maintained. This would be in line with findings reported by Bachman (1992) that a high degree of agreement can be obtained among raters when they are trained.

The frequency and magnitude of continually assessed work is another problem area especially at tertiary level. On the one hand, students want and need frequent feedback and plenty of opportunity

to practise but this may lead to the trivialisation of tasks. On the other hand, if tasks are to be authentic (a point which will be expanded later in this paper) and replicate the academic study cycle, then the number of tasks will automatically be limited. Language Centre courses in fact attempt to replicate the study cycle experienced by students in their own faculties. Hence there is a tendency to put great emphasis on project work which requires the students to define their 'problem', search for their own information, assimilate it from different sources and then present it in an acceptable academic form (either as an oral report or written presentation). This means that although students may receive guidance at intervening stages of the work, the opportunity of assessment may be all too limited.

Assessment: for whom?

It is often forgotten that parties other than the student and language teacher may be interested in students' assessment. Who those parties are will vary according to the academic level of the students and the reasons for the course they are taking. At HKU in addition to the student and language teacher one must also include the Language Centre as a teaching and administrative unit, the faculties and future employers. Each has their own specific needs which need to be addressed.

The Student For students, the primary role of assessment is feedback. If students are to make progress, they need extensive comments on both the positive and negative aspects of their work; what they have mastered and what they have failed to master. They also need to know what they could do to improve. This would imply that they do not in fact require a mark or grade, detailed comments should suffice. But students expect a grade. They like to know where they stand in relation to their colleagues and how good their piece of work is in the eyes of the teacher.

The Language Teacher Assessment for the language teacher has a retrospective as well as a prospective function. It allows teachers to reflect on course objectives the methodologies they have used, and to adapt accordingly. It also allows them to get to know their students -- to find out how much they know and how much they have learned during their course.

The Language Centre As an administrative unit, the Centre has to be in the position to demonstrate the effectiveness of its courses. Recently, the increased funding for English enhancement has added to the burden of accountability. Yet the limited time given over to language study (approximately 60 hours spread over two terms²) constrains the form this accountability can take. It would be unrealistic, for example, to try and show a gain in proficiency on an internationally recognised proficiency test. Accountability must in part take the form of subjective opinions collected from the participants in the teaching-learning process, through such means as questionnaires and interviews. However, such qualitative data needs to be supplemented as far as possible by quantitative data. Data of this type also provides invaluable input for the evaluation of courses, therefore monitoring of changes or gains in group performance needs to be built into the system of assessment.

One means used by the Language Centre to show group improvement is that of an oral test, with parallel forms being administered at the beginning and end of the English for Arts Students' course. The test simulates a tutorial (for more details see Morrison & Lee, 1985), assessed on a nine point (criterion-referenced) scale by two raters -- a tutor and a marker whose mark is double-weighted.

For the academic year 1991/92, of the 464 students who took both tests, 316 (68%) improved by at least one band while 114 (24.5%) showed a decrease of one band or more. Taking into account individual variability and notwithstanding the limitations of using bands as absolute marks of equal intervals for demonstrating gain (see Alderson, 1991b), the difference in mean scores was significant at the .001 level. The group as a whole performed better on the post test than the pretest. This may, of course, be a result of a number of factors, including students' familiarity with the test format as well as and with each other, their increased confidence having been at the University for nearly a year, and their general improvement in spoken English. Even though the gain cannot be attributed to any single factor, it is significant and likely at least in part to be due to the teaching of oral skills for tutorials and seminars on the Language Centre course.

Faculty Faculty needs are similar to those of the Language Centre. For the faculties grades are functional. They need to know which students have not satisfied the language requirement and what grade (if any) to put on the transcript. One area of difficulty that arises is that of comparability of grades and marks. Language teachers and subject specialists need to talk the same language in assessment and from the limited work so far undertaken in the area of joint assessment at HKU, it appears that not only do the two parties look for different things (justifiably enough), but that their interpretation of a marking scales may be very different. Much work remains to be done with regard to faculty and language teacher collaboration if students are not to get mixed messages from the different markers where, for example, a B+ for some teachers is 'average' while for others it is 'very good'.

The Employer One of the reasons why the university is taking an ever increasing interest in English is that employers are complaining about the poor communicative ability of graduates. Employers want some quality assurance. However, this is very difficult to give since for logistical reasons English enhancement courses are run during the first year of studies, in most cases two years before the students enter the employment market. Unless an assessment of students' English is undertaken in the final year, the grades given for performance in English may remain of little value, yet this is what the employer will go on. The university is still a long way from redressing this problem, however steps are being taken with the curriculum design to ensure that students are taught and assessed on some of the skills they will need later in life. The course to be piloted by the Language Centre for the Faculty of Engineering is 'English for Professional and Technical Communication' and not one restricted solely to English for Academic Purposes.

What should be Assessed?

Having ascertained that both continuous assessment and tests have a place in the curriculum, it was necessary to review the 'content' of assessment. Traditionally, the Language Centre has restricted continuous assessment to assessing achievement, but has extended testing beyond what has been specifically taught on the English courses to give a measure of proficiency. The end-of-course test for English for Arts Students, for example, is an integrated test of reading, listening and writing, even though the course objectives place much more emphasis on writing than the other two language skills. This could be regarded as unnecessary or even unfair. However, as explained above, an assessment of proficiency is required by some of the parties involved in the assessment and therefore appears to be justified.

Even within the sphere of continuous assessment it is difficult to determine what should be assessed. Unlike most faculty-based courses, English enhancement courses are designed to develop and strengthen skills rather than to teach content, though some content may be included. Metalinguistic and metacognitive skills (abilities to reflect on language and cognition) are among those given high priority. For example, much emphasis is placed on critical questioning in the Academic Communications and Study Skills course for Social Science students. Thus, tasks set for the students try to embody these skills, but such tasks cannot be devoid of content and the unresolved question that remains is whether the content should be 'authentic' in terms of what the students are studying or specific for the English enhancement course. The former increases the face validity of what is being assessed for the students; they can identify with the content. The latter, on the other hand, has the advantage of aligning the English courses more closely to the other courses students are studying in that it gives the English course its own content. It too has face validity but of a different kind; face validity for the faculty rather than for the students.

Another problem that needs further consideration is whether the product, process or effort involved in completing a task should be assessed. If the focus of teaching is on the process and revision of text is seen as a major contributor to the successful completion of a task, is it realistic or even fair to assess only the final product? Furthermore, how does one assess drafts? In real life drafts are often commented on by colleagues or one's boss and revised accordingly, but ultimately it is the final product that counts.

The Question of Authenticity

Authenticity is a key concern not only for continuous assessment but also for tests since it may affect the tasks students are required to complete. Nonetheless, it must be remembered that authenticity extends beyond tasks; texts selected as a basis for task completion may be 'authentic', as may the desired outcome of any given task. Each has to be considered and weighted against such other factors as the time taken for task completion, the costs involved in setting up the tasks as well as the generalizability from one task to another and the reliability of the measure. If, for example, students are required to complete an extended task such as a project which takes a large proportion of one term's teaching, there will be little time for other work.

A project may appear authentic in that it requires of the students a detailed academic investigation involving them in a complete study cycle. However, is this what is actually required of students in their faculties and will this be required of them in the future? It is likely that some of the skills involved in each task are relevant, but the task as a whole may be far from 'authentic' in its narrow sense of mirroring real-life outside the language classroom. (Faculties often require of the students considerably less than the Language Centre in terms of written and oral work, partly because of their belief that students are not able to cope with such high demands.) This does not necessarily invalidate the task. If one accepts that no task for assessment can replicate real-life, but each will have its own authenticity (Alderson, 1981), one needs to look for characteristics that overlap between the two contexts. To use Bachman's (1992) terminology, the tasks will have a varying degree of 'perceived relevance'. One therefore cannot look at tasks as being either authentic or inauthentic: authenticity should be viewed as a continuum.

Constraints of time and quantity of input are, of course, more severe in a test situation than for continuous assessment which in turn may affect task authenticity. This, however, may not be significant provided that there is authenticity of outcome, i.e. what the students have to produce has a high degree of authenticity in relation to the work they are expected to do for the subjects in which they are majoring. In other words, in a testing situation what appears to be important is not that the texts and tasks are highly authentic, but the outcome is, allowing generalizations to be made about students' abilities.

Assessment Criteria

It is not uncommon for assessment criteria to be predetermined for tests and examinations but left to individual teacher's judgements for continuous assessment. The latter may be a source of considerable variability as has been shown in a recent study by Williamson (this volume). The question therefore is whether it would be possible and indeed desirable to establish universal assessment criteria spanning continuous assessment as well as tests. In an ideal world, having one set of criteria that would be acceptable to all the parties interested in student assessment would be advantageous and, indeed, work is being carried out at the Language Centre³ and elsewhere (see North, 1992) to see if such criteria could be drawn up. However, as has been shown above, the needs and expectations of those involved in assessment are often very different and a number of factors including whether the criteria are to show achievement (for the student) or proficiency (for the future employer) have to be taken into account.

If assessment is to demonstrate achievement, should it be task-based and if so, how should a task be interpreted? A task may be as small as writing an introduction to an essay or as large as writing a project on, for example, the medium of instruction in Hong Kong. These 'tasks' are obviously not comparable and one could differentiate between them by looking on the former as an exercise while the latter as a task. This may solve one problem in that exercises could be used for feedback and not as part of formal assessment. But it does not solve the major problem of whether the product and/or the process of the tasks should be assessed. The larger the task, the more is involved in its completion and the more important is the process of completing it. Deriving criteria for assessing the product would appear a feasible proposition, but using the same for assessing the process may prove problematic or even counterproductive.

An additional source of concern is whether language or skills should be assessed. At the tertiary level where students have already undergone 1,000 - 1,500 hours of English language teaching at school,

the emphasis should be on developing skills and teaching students to use the language they have efficiently. But it is very difficult to predetermine which skills need to be utilised for the completion of a task or how to judge a student who has completed a task successfully, but on calling upon different skills to those being assessed. Furthermore, is it possible to weight these skills in any purposeful way and to what extent can one make generalizations about students' mastery of specific skills on their performance on any one task?

The detail with which the criteria need to be specified may also differ according to the purpose of assessment. As suggested above, for assessing achievement the criteria could be task specific which would be most beneficial for the students but of little value to faculty or future employers. The latter would want specifications they could relate to; they may even want to know where one individual lay not in relation to his/her group but in relation to the whole population. In other words, they may need to know where the student lies on a general proficiency scale such as IELTS. A possible solution that was put forward was to work out a two tier system, the first being a more global one that relates performance to an accepted proficiency scale and the second a more detailed one that relates to achievement. Since all students at HKU have attained a minimum of a grade D in the Use of English examination, the more detailed descriptors would in effect spread the students who would otherwise fall within a rather narrow band on a proficiency scale. However, one must bear in mind that the more detailed the specifications and the finer the distinctions being made, the more difficult the criteria are to apply.

'Alternative Assessment'

One view expressed was that current assessment procedures are seen as threatening. A great deal of emphasis is placed on a number of major assignments and the end-of-course test. Furthermore, there is little flexibility built into the system to allow students to progress at the pace they would feel most comfortable with or to actively participate in setting their own goals. Greater student involvement in assessment would go some way in alleviating pressures hitherto experienced. Students could, for example, be participants in building up their learning profiles with their teachers. They could also set their own agenda for assessment within their teacher's framework. And if they were taught to assess themselves and to take a greater responsibility for their own learning, they would begin to understand the assessment process and would, hopefully, no longer see it as threatening. Assessment would become a motivating factor that could enhance performance.

Conclusion

From the discussions it appears that both testing and continuous assessment have a place in the tertiary curriculum, although more could be done to make both less threatening and more accessible to the students. The Language Centre is moving in this direction: as of September 1992, it is planning to make available test marking criteria to students on two of its major courses to help students set their own learning goals; this is widely seen as a step in the right direction.

As in all assessment situations there are a number of tensions. There is a tension between demonstrating achievement and proficiency; there is also a tension between maintaining high authenticity of assessment procedures and efficiency so that assessment does not consume too much teaching time; and finally, there is the tension between allowing flexibility in the system while maintaining reliability of results. Since it would be impossible to remove these tensions, one needs to be aware of them and to preserve a balance between their conflicting demands.

Reliability of assessment is becoming an increasing concern as the demand for accountability grows. As courses become part of the degree curricula so students' performance needs to be recorded in a meaningful and comprehensible way. There is a growing need for external comparability as well as recognition throughout the University. There is also a need for a system that is fair to students and one that motivates them to do well. There is, in other words, a need for assessment to become an integral part of course development with all parties contributing to it, rather than its design being left to a small group.

Notes

¹ The Language Centre was responsible for the teaching and assessment of English to first year undergraduate students not majoring in English at the time of writing. However, as of 1 July, 1992, this operation has been transferred to a separate unit, the English Centre.

² The time allocated to English enhancement courses varies across faculties. The course for the Faculty of Arts is 60 hrs, 12 of which are for self-access work and small group tutorials while the course for students of the Faculty of Engineering is 48 timetabled hours and in addition students are expected to undertake self-access work.

³ This work was started as an attempt to reach a common understanding about marking criteria among staff of the Language Centre and the Faculty of Social Science. It was initiated by Nigel Bruce who should be contacted for more details on the project.

Acknowledgements

My thanks go to all members of the Language Centre who attended this workshop. Without them this paper would not have been possible. My thanks also go to Lily Leung for drawing up and letting me use the information in table 1 and to Desmond Allison for his comments and suggestions on the initial draft.

References

- Alderson, J. C. (1981). Report of the discussion on communicative language testing. In J. C. Alderson & A. Hughes (Eds.), Issues in Language Testing *ELT Documents No. 111* 55-65. London: The British Council.
- Alderson J. C. (1991a). Language testing in the 1990s: How far have we come? How much further have we to go? In S. Anivan (Ed.), Current Developments in Language Testing 1-26. Singapore: SEAMEO.
- Alderson, J. C. (1991b). Bands and scores. In J. C. Alderson and B. North (Ed.) Language Testing in the 1990s 71-86. London: Macmillan Publishers.
- Bachman, L. (1991). What does language testing have to offer? TESOL Quarterly, 25/4 671-704.
- Morrison D. M. & Lee, N (1985). Simulating an academic tutorial: A test validation study. In Y. P. Lee, A. Fok, C. Y. Y. Lord & G. Low (Eds.) New Directions in Language Testing 127-135. Oxford: Pergamon Press.
- North, B (1992). Options for scales of proficiency for a European language framework. NFLC Occasional Paper 9-21.
- Williamson, D. (1992). A Progress Report of an Action Research Project into the Marking of Students' Written Work. Hongkong Papers in Linguistics and Language Teaching, 15.